Proposta A: Calcular o desequilíbrio de ligação entre dois alelos em uma determinada população.

A associação não aleatória de alelos localizados em diferentes loci é chamado de desequilíbrio na fase gamética ou desequilíbrio de ligação (LD)¹.

Em outras palavras o desequilíbrio de ligação ocorre quando os genótipos em dois loci não são independentes um do outro.

O coeficiente de desequilíbrio de ligação (D) é dado pela diferença entre o produto das frequências dos haplótipos² em acoplamento e o produto das frequências dos haplótipos em repulsão.

Considere dois loci (A e B), cada um segregando dois alelos (A1, A2, B1 e B2). Existem quatro haplótipos possíveis presentes na população:

|  |  |
| --- | --- |
| Haplótipo | Frequência |
| A1B1 | X11 |
| A1B2 | X12 |
| A2B1 | X21 |
| A2B2 | X22 |

A frequência dos alelos pode ser expressa em relação a frequência dos gametas:

|  |  |
| --- | --- |
| Alelo | Frequência |
| A1 | p1 = X11 + X12 |
| A2 | p2 = X21 + X22 |
| B1 | q1 = X11 + X21 |
| B2 | q2 = X12 + X22 |

Se os alelos nos dois loci estão aleatoriamente associados um com o outro, então as frequências dos quatro haplótipos será igual ao produto das frequências de seus alelos:

X11 = p1q1

X12 = p1q2

X21 = p2q1

X22 = p2q2

Nesse caso não há desequilíbrio de ligação e as frequências dos haplótipos pode ser calculada a partir da frequência dos alelos naquela população.

Se os alelos nos dois loci não estão aleatoriamente associados, então haverá um desvio (D) nas frequências esperadas:

X11 = p1q1 + D

X12 = p1q2 - D

X21 = p2q1 - D

X22 = p2q2 + D

O parâmetro D é o coeficiente de desequilíbrio de ligação, proposto por Lewontin e Kojima³.

D = X11X22 – X12X21

Aonde, X11 e X22 são os haplótipos em acoplamento e X12 e X21 os haplótipos em repulsão.

Os valores de D variam entre -0,25 a 0,25

O valor máximo de D muda em função das frequências alélicas nos dois loci, Lewontin4 então propôs a normalização de D em relação ao seu máximo valor possível.

D’ = D/Dmax

Quando D > 0

Dmax é igual ao menor valor de p1q2 ou p2q1

Quando D < 0

Dmax é igual ao menor valor de p1q1 ou p2q2

D’ varia entre 0 e 1

Outra forma de se calcular o desequilíbrio de ligação é através da correlação entre pares de loci5.

r² = D² /p1p2q1q2

Se as frequências dos alelos forem iguais, então r² varia entre 0 e 1.

Quando r² = 1, os loci estão em completo desequilíbrio de ligação.

A função irá calcular os valores de D’, r² , a frequência estimada dos haplótipos, a frequência esperada em caso de equilíbrio de Hardy-Weinberg e o resultado do teste qui-quadrado.

Os dados de entrada serão do tipo data.frame e poderão ser lidos a partir de um arquivo txt.

O cálculo de LD será feito de forma pareada, o usuário poderá selecionar os alelos de interesse ou fazer a análise para todos os alelos do arquivo. Também será possível selecionar os alelos selecionando a região genômica, nesse caso o data.frame deverá conter informações sobre a posição física (bp) dos alelos.

Os dados de saída estão na forma matriz contendo os valores de D’, r² e χ² para cada par de alelos testados.

Os resultados da frequência estimada dos haplótipos e da frequência esperada serão visualizados apenas quando o usuário escolher apenas dois alelos o cálculo do LD. Nesse caso o resultado será um data.frame.

Proposta B: Determinação de modelos de regressão logística para fenótipos complexos.

A função irá determinar quais as variáveis estão relacionadas com o fenótipo. Todas as possíveis variáveis preditoras serão testadas, e serão selecionadas as que apresentarem modelo com p-value ≥0.05.

As variáveis selecionadas serão submetidas a um teste para identificar colinearidade entre elas.

Por fim será construído um modelo de regressão logística com base nas variáveis selecionadas.

Os dados de entrada serão em forma de data.frame a partir de um arquivo txt e os dados de saída serão um data.frame contendo o modelo de regressão que melhor explica o fenótipo em questão e um gráfico com os dados e o modelo de regressão.